DOCUMENT RESUME

ED 461 382 IR 056 397

AUTHOR Lunau, Carrol; Turner, Fay
TITLE Issues Related to the Use of Z39.50 To Emulate a Centralized Union Catalogue.
INSTITUTION National Library of Canada, Ottawa (Ontario).
SPONS AGENCY Association of Research Libraries, Washington, DC.
PUB DATE 1997-04-00
NOTE 8p.; Prepared for the ARL Access Committee.
AVAILABLE FROM For full text: http://www.nlc-bnc.ca/resource/vcuc/ezarl2.htm.
PUB TYPE Reports - Evaluative (142)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Access to Information; *Bibliographic Databases; Bibliographic Records; Cataloging; Computer Networks; Foreign Countries; *Information Retrieval; Interlibrary Loans; Libraries; Library Circulation; Library Collections; *Library Cooperation; Library Networks; National Libraries; Online Catalogs; Online Searching; Search Strategies; *Shared Library Resources; *Union Catalogs; Vendors
IDENTIFIERS Barriers to Implementation; Canada; *Standardization; *Z39 50 Protocol

ABSTRACT

The virtual Canadian union catalog (vCuc) project is a project among Canadian libraries to use Z39.50 (an information protocol standard) for searching distributed individual library catalogs and union catalogs. This paper summarizes the major technical, vendor, bibliographic data, and administrative issues that must be resolved before the distributed catalog environment will be able to provide the functionality which is available with a centralized union catalog. Solutions and the parties who are responsible for implementing them (librarians, vendors, national libraries and library organizations, library customers, or a combination) are identified. Issues discussed include: retrieval of locations and holdings information; the variety of attributes supported by vendors; inconsistency in vendor and library mapping of use attributes; ability to link from client system to other applications; merging result records of multiple systems; non-standardized use of library symbols; searches for specific serials issues; item level data and need for linkage of circulation and bibliographic records; incomplete cataloging; search terms supported by library; placement of location information and ease of user retrieval; description of Z39.50 features and options available in target system; and charging for the service. Vendors, technical developers, and librarians must continue to work together to find solutions which will assist libraries in providing service to their clients. (SWC)

National Library of Canada | Bibliothèque nationale du Canada



# Issues Related to the Use of Z39.50 to Emulate a Centralized Union Catalogue

Prepared for the ARL Access Committee
by
Carrol Lunau & Fay Turner
National Library of Canada
April 1997



## Table of Contents





The virtual Canadian union catalogue (vCuc) project is a project amongst Canadian libraries to use Z39.50 for searching distributed individual library catalogues and union catalogues. Preliminary results indicate that there are a number of technical, bibliographic, vendor and administrative issues which must be resolved before the distributed catalogue environment will be able to approach providing the functionality which is available with a centralized union catalogue. Many of these issues have been identified in the recent article by Clifford Lynch.[1]

This discussion document provides a brief summary of the major issues and, where appropriate, proposes a mechanism for developing a solution.

## 1. Technical Issues

### 1.1 Locations and holdings

The retrieval of locations and holdings information is central to a distributed union catalogue. There are several options available for the provision of location and holdings information in response to a Z39.50 query including the use of the MARC bibliographic record, the MARC Holdings record and

the Z39.50 OPAC Record. At the present time there is no consistency in the use of these options, not even in the use of the MARC tags within the MARC bibliographic record. Agreement is required among implementors on a consistent method for the supply of location information, summary holdings information, detailed holdings information and circulation information. This is a very complex, technical and developmental issue.

**Responsibility**

The National Library of Canada has prepared a discussion paper which was discussed at the April 1997 ZIG (Z39.50 Implementors Group) meeting.

The ZIG is working towards a solution for the transfer of location, holdings and circulation data.

Vendors must also develop the links between the circulation availability data and the bibliographic records which are available through the Z39.50 server.

# 2. Vendor Issues

## 2.1 Attributes Supported by Vendors

Target databases and servers support a variety of Z39.50 Attributes (search terms) and attribute combinations. In addition, different implementors attach different semantics to the same attributes. When searching across multiple servers the search must conform to the lowest common denominator in order to be successful at each site.

See also 3.2

**Responsibility**

This problem is widely acknowledged within the Z39.50 community. Vendors need to agree on a minimum set of Use attributes which go beyond the basic Author, Title, and Subject to include control numbers such as ISSN and ISBN. However, this may require vendors to re-index their databases which can be costly. In addition, there needs to be an agreement on the semantics of all attributes.[2] Libraries need to encourage vendors to implement a common subset of attributes and attribute combinations. This is a shared responsibility between libraries and vendors

## 2.2 Mapping of Use Attributes

Z39.50 servers map the Use attributes to the indexes in the database for the records. Some vendors have been very broad in their mapping of the attributes so that a search may return records that are meaningless. For example, a search using the Author attribute and author's name may return records with the author's name contained in a title, a conference name, subject, notes, etc. This type of imprecision in the records returned can be frustrating for the end user.

See also 3.3

**Responsibility**

This is a well -known problem with in the Z39.50 community. The ZIG has encouraged vendors to

return a diagnostic message of "Unsupported Attribute Type" rather than imprecise results. Vendors are responsible for solving but the library community should also encourage the vendors to be more precise in their mapping of Z39.50 attributes.

## 2.3 Links from Client System to Other Applications

The search of a distributed union catalogue is conducted within the context of another operation such as ILL, reference, cataloguing, scholarly research, etc. The data which is retrieved is generally needed for use in a bibliography, an ILL request or some other application. Therefore, Z39.50 client systems must be able to transfer the data seamlessly into a subsequent application which may also be protocol-based. For example, an integrated ILL workstation would include both Z39.50 and ISO 10160/10161.

Client system vendors need to be encouraged to expand on the capabilities of their clients so that data retrieved via Z39.50 can be integrated with other applications such as ILL protocol- based systems.

**Responsibility**

The National Library will continue to work with vendors of library software to explain the development of the vCuc, the evolution of the environment from centralized to distributed catalogues, and the need for links from client systems to other applications. Since the library automation vendors are the same vendors in Canada and the US, ARL and NLC can cooperate on working with the vendors.

## 2.4 Merging Result Records

When searching multiple systems simultaneously the user may receive many duplicate records. This can occur in an environment where a search includes a centralized database as well as individual library catalogues whose holdings are also included in the centralized database. Also, a number of libraries may have the same title and the searcher receives multiple records for the same bibliographic item.

**Responsibility**

Ideally, client systems would perform some level of matching and amalgamation of search results from disparate systems. On the simplest level this matching could be based on one of the control numbers such as ISSN or ISBN, however sophisticated matching algorithms would be needed to produce the levels of record amalgamation which are now common in large union catalogue databases such as OCLC or AMICUS. Implementation of these types of complex algorithms is not possible due to the prohibitive size of the application, the costs and the detrimental impact on response time; nonetheless, vendors need to recognize that merging of records is an important requirement for decentralized union catalogue activity and search for realistic solutions. This requirement for merged result sets, in the context of multi-database searching, has been identified by Cornell University's Albert R. Mann Library user study for the design of the next generation "Gateway".

Action: Vendors; ARL and NLC to encourage the development of this feature in vendor clients.

## 2.5 Library Symbols

See 4.2 below.

### 2.6 Searches for Specific Serials Issues

Within ILL, serial requests are generally requests for a specific article in a specific issue, e.g.. v.10, no. 2. The Z39.50 standard has been extended to accommodate the specification of a query for a specific issue by using the SICI. However, this extension has not been implemented by vendors so the user often has to scan several screens of holdings information to locate the required specific issue required. This problem is aggravated with union catalogue records which can include holdings from many libraries.

**Responsibility**

Action: Vendors

### 2.7 Item Level Data

Users often want to know if a specific item is available for them to borrow. In order to know this the circulation files and the bibliographic records must be linked.

See also 1.1

**Responsibility**

Action: Vendors

## 3. Bibliographic Data Issues

### 3.1 Incomplete Cataloguing

While the target may support a rich set of attributes, the databases themselves may not contain key indexes, such as those for standard numbers like ISSN. For example, a search against 6 Canadian databases for the serial American Historical Review resulted in the following when searched under:

title --56 hits but only 35 valid
title + ISSN --28 hits but only 25 valid
ISSN --24 hits and all 24 valid

In other words, by using title 37.5% of the records retrieved were completely irrelevant to the search but when using the most precise search key, ISSN, 11 valid records were not retrieved. The valid records were not retrieved because 2 of the servers did not support ISSN and 10 of the bibliographic records did not include the ISSN. This example is a relatively precise search and the number of false hits escalates dramatically with less precise searches.[3]

**Responsibility**

Libraries must include control numbers such as ISSN and ISBN in their cataloguing records in order to increase the precision of search results with multiple server searching.

NLC to continue to publicize the need to include the data in bibliographic records.

## 3.2 Attributes Supported by Library

Some Z39.50 servers are configurable to allow the customer to select which of the attributes supported by the server will actually be used., i.e. turned on. Therefore, even though the server allows searching on numerous search terms, the request may be unsuccessful because the terms have been disallowed by the library.

**Responsibility**

One solution is for libraries to agree to a minimum set of Use Attributes which goes beyond the basic author, title and subject to include control numbers such as ISSN and ISBN.

NLC & vCuc partners will consider developing an agreement for the minimum set of Attributes.

## 3.3 Library Mapping of Use Attributes

Where the library customer has control over the mapping of Z39.50 use attributes to database indexes and/or MARC fields, often the indexes and fields selected are semantically much broader than the attribute term itself. The search results are therefore often imprecise and unexpected. For example: name may be mapped to several MARC fields such as 100, 110, 111, 400, 410, 411, 508, 511, 600, 611, 692, 693, 694, 700, 705, 710, 711, 715, 765, 767, 780, 785, 787, 800, 810, 811, 870, 872 by one implementor and to different MARC tags by another. A user looking for a novel written by a specific author may receive a number of false hits because of this mapping.

Information about the mapping of data between the local database and the Z39.50 target is not generally available but the searcher may need to understand this mapping in order to formulate an effective search statement or to interpret the search results

See also Issue 2.2

**Responsibility**

More precise mapping is required between the use attribute and the database indexes. When a use attribute is not supported, a diagnostic message to that effect should be returned. Libraries that can determine the mapping within their own systems should be aware of the negative impact resulting from the application of use attributes to a wide number of indexes and MARC fields.

# 4. Administrative Issues

## 4.1 Placement of Location Information

When searching large union catalogues the user must be able to easily determine which library actually holds the desired item. Current practice varies, both in terms of the inclusion of the library symbol and in the placement of the data in the bibliographic record. For example: some libraries include a symbol with the local call number but others include it with the holdings information.

**Responsibility**

Libraries need to agree on the use of a common field for the location information and the inclusion of the library symbol.

NLC will work with vCuc partners to resolve this.

## 4.2 Library Symbols

Library symbols used within records can be problematic when bibliographic utilities or systems within a consortium utilize local symbols which are known only to their membership. A searcher from outside the local consortium may not be familiar with the local symbols and may not have access to a local directory containing the information. As well, some libraries are planning to make Z39.50 clients available to their patrons and patrons will not be able to interpret symbols used in various systems.

**Responsibility**

Libraries in North America need to agree on a standard use of symbols. Within Canada, NLC encourages the use of The Symbols of Canadian Libraries assigned by the National Library of Canada. As well, the 2 character country code should be included to avoid duplication of symbols which may occur between countries -e.g., CaOONL for the National Library of Canada.

The standardization of symbols has been discussed by the Conference of Directors of National Libraries and is being worked on by ISO TC46.

## 4.3 Description of Z39.50 Targets

To improve search results, it is important for the client system and even the end user to have information about the Z39.50 features and options implemented by the target system. The Z39.50 standard contains an EXPLAIN service which consists of a searchable database containing information about the Z39.50 target system. Ideally, a client system would search the EXPLAIN database of a remote target system and configure itself for that system. However, because EXPLAIN has not yet been widely implemented, libraries and software vendors need to describe their systems on the Web using a standardized set of information fields.

**Responsibility**

Information about remote targets is essential. To assist in the description of these systems NLC drafted a template for Web pages. For the vCuc project, a Directory of Z39.50 Targets in Canada has been created. Other sites, such as the one maintained by Sirsi also exist to provide this information. LC, OCLC, and RLG also have Web pages describing their Z39.50 implementations.

However, this information is not available for most Z39.50 implementations and installations. A coordinated approach with the vendors and the library sites to provide server information would greatly improve Z39.50 search results.

NLC to promote the development and use of a directory of Canadian Z39.50 targets. The promotion of a North American directory could be included in NAILDD discussions of directories.

### 4.4 Charging

Charging can become complex in this environment. Charging practices have traditionally been different for searching a database than for downloading a record for copy cataloguing. Within Z39.50 the lines become blurred since all records are "downloaded" when returned by the server. Also, a high percentage of returned records are invalid hits for which the user should not be charged.

**Responsibility**

Database providers will need to develop a charging algorithm which can accommodate these factors within the Z39.50 environment.

Action: Information providers.

## 5. Conclusion

This document provides a brief thumbnail sketch of some of the library and technical questions which arise when using Z39.50 to emulate a distributed union catalogue. Vendors, technical developers and librarians must continue to work together to find solutions which will assist libraries in providing service to their clients.

## 6. References

1. Lynch, Clifford A., "Building the infrastructure of resource sharing: union catalogs, distributed search, and cross-database linkage".*Library Trends*, 45 (3), winter 1997. p.448-461

2. Lynch, Clifford A., "The Z39.50 information retrieval standard. Part 1: A Strategic view of its past, present and future." *D-Lib Magazine*, April 1997

3. Dorman, David, "Technically speaking: new developments shake up midwinter exhibits". *American Libraries*, v.28, no.4, April 1997, p.78

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

# NOTICE

## REPRODUCTION BASIS

This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☐

This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").